A SYNTHETIC DNA ENCODING AN ORANGE SEAPEN-DERIVED GREEN FLUORESCENT PROTEIN WITH CODON PREFERENCE OF MAMMALIAN EXPRESSION SYSTEMS AND BIOSENSORS

BENEFIT OF PRIOR PROVISIONAL APPLICATION

This utility patent application claims the benefit of co-pending prior U.S. Provisional Patent Application Serial No. 60/297,645, filed June 12, 2001, entitled "Synthetic DNA Encoding A Green Fluorescent Protein With The Most Favored Codons For Mammalian Systems" having the same named applicants as inventors, namely, Yih-Tai Chen and Longguang Cao.

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to an isolated and purified DNA encoding a humanized bioluminescent green fluorescent protein (hPtFP) derived from the orange seapen *Ptilosarcus gurneyi* in which all the codons are the favored or most favored codons for mammalian expression systems. Truncation mutants of the humanized *Ptilosarcus gurneyi* fluorescent protein (hPtFP) of the present invention are functional as fluorescent reporter molecules in a biosensor system. The green fluorescent protein of the present invention is useful as an improved fusion partner in cellular proteins allowing direct observation of the behavior of the tagged protein.

2. Description of the Background Art

A major component of the new drug discovery paradigm is a continually growing family of fluorescent and luminescent reagents that are used to measure the temporal and spatial distribution, content, and activity of intracellular ions, metabolites, macromolecules and organelles. Classes of these reagents include labeling reagents that measure the distribution and amount of molecules in living or fixed cells, environmental indicators to report signal transduction events in time and space, and fluorescent protein biosensors to measure target molecular activities within living cells. A multiparameter

approach that combines several reagents in a single cell is a powerful new tool for drug discovery.

Those skilled in this art will recognize a wide variety of fluorescent reporter molecules that can be used in the field of drug discovery. Particularly, herein are disclosed novel humanized fluorescent proteins. Similarly, fluorescent reagents specifically synthesized with particular chemical properties of binding or association have been used as fluorescent reporter molecules. (Barak et al., (1997), J. Biol. Chem. 272:27497-27500; Southwick et al., (1990), Cytometry 11:418-430; Tsien (1989) in Methods in Cell Biology, Vol. 29 Taylor and Wang (eds.), pp.127-156). Fluorescently labeled antibodies are particularly useful reporter molecules due to their high degree of specificity for attaching to a single molecular target in a mixture of molecules as complex as a cell or tissue. However, fluorescently labeled antibodies present several limitations.

It is known that luminescent probes can be synthesized within the living cell or can be transported into the cell via several non-mechanical modes including diffusion, facilitated or active transport, signal-sequence-mediated transport, and endocytic or pinocytic uptake. Mechanical bulk loading methods, which are well known in the art, can also be used to load luminescent probes into living cells. (Barber et al. (1996), Neuroscience letters 207:17-20; Bright et al. (1996), Cytometry 24:226-233; McNeil (1989) in Methods in Cell Biology, Vol. 29, Taylor and Wang (eds.) pp. 153-173). These methods include electroporation and other mechanical methods such as scrape-loading, bead-loading, impact loading, syringe-loading, hypertonic and hypotonic loading. Additionally, cells can be genetically engineered to express reporter molecules such as Green Fluorescent Protein, coupled to a protein of interest as previously described (Chalfie and Prasher U.S. Patent 5,491,084; Cubitt et al. (1995), Trends in Biochemical Science, 20:448-455).

Luminescence is the process whereby a molecule is electronically excited and releases light when it returns to a lower energy state. Bioluminescence is the process by which living organisms emit light that is visible to other organisms. In bioluminescence the excited state is created by an enzyme-catalyzed reaction. The color of the emitted light in a bioluminescent reaction is characteristic of the excited molecule, and is independent from its source of excitation and temperature.

Molecular oxygen is known to be essential in some well characterized bioluminescent systems, such as the bioluminescence of luciferase. Luciferases are oxygenases, that act on a substrate, luciferin, in the presence of molecular oxygen and transform the substrate to an excited state. Upon return to a lower energy level, energy is released in the form of light. Ward et al., Chapter 7 in Chemi- and Bio-luminescence, Burr ed. Marcel Dekker, Inc. NY, pp. 321-358; Hastings, J.W. (1995) Cell Physiology: Source Book, N. Sperelakis (ed.), Academic Press, pp. 665-681; Luminescence, Narcosis and Life in the Deep Sea, Johnson Vantage Press, NY, pp. 50-56. Bioluminescent species span many genera and include microscopic organisms, including bacteria, primarily marine bacteria such as *Vibrio* species, fungi, algae, and dinoflagellates, to marine organisms including arthropods, mollusks, echinoderms, and chordates, and terrestrial organisms including annelids and insects.

Luminescence (bioluminescence, chemiluminescence, and fluorescence) is used for qualitative and quantitative determination of specific substances and processes in biology and medicine. For example, various luciferase genes from various organisms have been cloned and exploited as reporters in numerous assays. On the other hand, treating cells with dyes and fluorescent biomolecules allowing imaging of the cells, and genetic engineering of cells to produce fluorescent proteins as reporter molecules are useful detection methods known by those persons skilled in the art. For instance, treating cells with dyes and fluorescent biomolecules allowing imaging the cells, and genetic engineering of cells to produce fluorescent proteins as reporter molecules are useful detection methods known in the art. Wang et al., Methods in Cell Biology, New York, Alan R. Liss, 29:1-12, 1989. One such fluorescent reporter protein is the green fluorescent protein (GFP) of the jellyfish *Aequorea victoria* which absorbs blue light with an excitation maximum at 395 nm, with a minor peak at 470 nm, and emits green fluorescence with an emission maximum at 510 nm, with a minor peak near 540 nm and does not require an exogenous factor. However, the absorption and emission spectra for *Aequorea* GFP present certain limitations. The excitation and emission maxima of the wild type *Aequorea* GFP are not within the optimal range of wavelengths of standard fluorescence optics.

The green fluorescent proteins (GFP) constitute a class of chromoproteins found among certain bioluminescent coelenterates. These proteins are fluorescent and function as the ultimate bioluminescence emitter in these organisms by accepting energy from enzyme-bound, excited state oxyluciferin. *Ward et al.*, (1982) Biochemistry 21: 4535-4540.

Uses of *Aequora* GFP for the study of gene expression and protein localization are discussed in Chalfie et al., Science 263:802-805, 1994. Some properties of wild-type *Aequora* GFP are disclosed by Morise et al., Biochemistry 13:2656-2662, 1974, and Ward et al., Photochem. Photobiol. 31:611-615, 1980. An article by Rizzuto et al., Nature 358:325-327, 1992, discusses the use of wild-type *Aequora* GFP as a tool for visualizing subcellular organelles in cells. Kaether and Gerdes, FEBS Letters 369:267-271, 1995, report the visualization of protein transport along the secretory pathway using wild-type *Aequora* GFP. The expression of *Aequora* GFP in plant cells is discussed by Hu and Cheng, FEBS Letters 369:331-334, 1995, while *Aequora* GFP expression in *Drosophila* embryos is described by Davis et al., Dev. Biology 170:726-729, 1995.

U.S. Pat. No. 5,491,084 discloses expression of GFP from *Aequorea victoria* in cells for use as a reporter molecule fused to another protein of interest. PCT/DK96/00052 relates to methods of detecting biologically active substances affecting intracellular processes by utilizing a GFP construct having a protein kinase activation site. GFP proteins are used in various biological systems. For example, PCT/US95/10165 describes a system for isolating cells of interest utilizing the expression of a GFP-like protein. PCT/GB96/00481 describes the expression of GFP in plants. PCT/US95/01425 describes modified GFP protein expressed in transformed organisms to detect mutagenesis. Mutants of GFP have been prepared and used in several biological systems. (Hasselhoff et al., Proc. Natl. Acad. Sci. 94:2122-2127, 1997; Brejc et al., Proc. Natl. Acad Sci. 94:2306-2311, 1997; Cheng et al., Nature Biotech. 14:606-609, 1996; Heim and Tsien, Curr. Biol. 6:178-192, 1996; Ehrig et al., FEBS Letters 367:163-166, 1995). Methods describing assays and compositions for detecting and evaluating the intracellular transduction of an extracellular signal using recombinant cells that express cell surface receptors and contain reporter gene constructs that include transcriptional

regulatory elements that are responsive to the activity of cell surface receptors are disclosed in U.S. Pat. No. 5,436,128 and U.S. Pat. No. 5,401,629.

Certain types of cells within an organism may contain components that can be specifically labeled that may not occur in other cell types. For example, epithelial cells often contain polarized membrane components. That is, these cells asymmetrically distribute macromolecules along their plasma membrane. Connective or supporting tissue cells often contain granules in which are trapped molecules specific to that cell type (e.g. heparin, histamine, serotonin, etc.). Skeletal muscle cells contain a sarcoplasmic reticulum, a specialized organelle whose function is to regulate the concentration of calcium ions within the cell cytoplasm. Many nervous tissue cells contain secretory granules and vesicles in which are trapped neurohormones or neurotransmitters. Therefore, fluorescent molecules can be designed to label not only specific components within specific cells, but also specific cells within a population of mixed cell types.

Those skilled in the art will recognize a wide variety of ways to measure fluorescence. For example, some fluorescent reporter molecules exhibit a change in excitation or emission spectra, some exhibit resonance energy transfer where one fluorescent reporter loses fluorescence, while a second gains in fluorescence, some exhibit a loss (quenching) or appearance of fluorescence, while some report rotational movements. (Giuliano et al. (1995), Ann. Rev. of Biophysics and Biomol. Structure 24:405-434; Giuliano et al. (1995), Methods in Neuroscience 27:1-16). The GFPs exhibit absorption at a particular wavelength, and emission at a different wavelength characteristic for each green fluorescent protein which sometimes allows for the pairing of GFP's with two distinct signals being detectable.

In addition to the limitations in detection with standard fluorescence optics presented by the absorption-emission wavelength spectrum of *Aequora* GFP, another difficulty is the potentially low level of fluorescent signal emitted by GFP transfected into a heterologous cell type. This is the result of low level expression normally associated with the expression of a non-native species protein being expressed by a cell, in this case a jellyfish protein being expressed in higher level organisms such as mammals. This is partly due to different codon usage in the native marine organism

sequences that are different from the host or transfected cell's codon usage. In spite of this background art, there remains a very real and substantial need for a fluorescent reporter molecule having a narrower absorption-emission wavelength spectrum and having an optimized expression in a host or transfected cell resulting in fluorescent signals that are easily detected with standard fluorescence optics.

U.S. Patent Nos. 5,786,464 (Seed et al.) and 5,795,737 (Seed et al.) dislcose replacing non-preferred codons with preferred codons to increase expression in mammalian cell lines of other proteins, such as the green fluorescent protein of the jellyfish *Aequorea victoria*.

U.S. Patent No. 5,874,304 (Zolotukhin et al.) discloses a humanized green fluorescent protein gene adapted from the jellyfish *Aequorea victoria*. U.S. Patent No. 5,968,750 (Zolotukhin et al.) discloses a method of labeling a mammalian cell comprising expressing a humanized green fluorescent protein gene in the cell wherein the genes have an increased number of GCC or GCT alanine-encoding codons in comparison to the wild type jellyfish gene sequence.

U.S. Patent No. 6,232,107 (Bryan et al.) discloses isolated and purified nucleic acids encoding green fluorescent proteins from the genus *Renilla* and *Ptilosarcus* and the green fluorescent proteins encoded thereby.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows comparative fluorescence of COS1 cells expressing the synthetic DNA versus a commercially available nuclear dye (Hoechst 33342 stain).

Figure 2 shows *in situ* fluorescence of the humanized *Ptilosarcus gurneyi* fluorescent protein in COS1 cells.

Figure 3 is a histogram of the fluorescent intensity in COS1 cells transiently transfected with wild type *Ptilosarcus gurneyi* green fluorescent protein DNA or the synthetic green fluorescent protein DNA. The X-axis of Figure 3 is the fluorescence intensity, minimum is zero and maximum is 4095. The Y-axis of Figure 3 is the normalized distribution (percentage) of the cell population.

Figure 4 shows two examples of the stable cell lines established by transfection with hPtFP green fluorescent protein DNA of the present invention. The left panel shows

stable A549 cells transfectants expressing hPtFP. The right panel shows stable HEK293 cell transfectants expressing hPtFP.

Figure 5 shows the non toxic effect of hPtFP on its mammalian host cells.

Figure 6, Panel A, shows COS1 cells transiently transfected with human CD7 and were stained with a monoclonal antibody against human CD7. Panels B & C show COS1 cells transiently transfected with CD7-fluorescent protein.

Figures 7 and 8 show a diagram of the configuration of the Caspase 3 and Caspase 8 biosensors, respectively.

Figure 9 shows the wild type *Ptilosarcus gurneyi* nucleotide sequence (top row) compared to the nucleotide sequence encoding the humanized *Ptilosarcus gurneyi* fluorescent protein of the present invention (bottom row).

Figure 10 shows the amino acid sequence (top row) and the double stranded nucleotide sequence (bottom rows) of the humanized *Ptilosarcus gurneyi* fluorescent protein of the present invention.

Figure 11 shows the full length nucleotide of humanized *Ptilosarcus gurneyi* fluorescent protein of the present invention including regions upstream and downstream to the coding region.

Figure 12 shows the full length protein sequence of humanized *Ptilosarcus gurneyi* fluorescent protein of the present invention from start codon to stop codon.

Figure 13 shows the (truncated) deletion mutants of hPtFP and their effects on green fluorescence.

Figure 14 shows HeLa cells transfected with the hPtFP-Caspase-8 biosensor of this invention before and after treatment with staurosporine.

Figure 15 shows a codon usage table for a human system, compiled from 22747 coding regions CDS's (10965560 codons).

Figure 16 shows a restriction endonuclease cleavage map for expression vector M2.

Figure 17 shows the nucleotide sequence of expression vector M2.

Figure 18 shows a general description of gene synthesis.

SUMMARY OF THE INVENTION

The present invention has met the hereinbefore described needs. The present invention provides an isolated and purified DNA encoding a green fluorescent protein from the orange seapen *Ptilosarcus gurneyi* in which all of the codons are favored for mammalian systems. The full length encoded protein of the present invention has 239 amino acid residues. Preferably, the encoded protein of the present invention is truncated having 224 amino acid residues and most preferably has 219 amino acid residues. In comparison to the wild type *Ptilosarcus gurneyi* green fluorescent protein having 238 amino acid residues, codons for 145 amino acids of the humanized *Ptilosarcus gurneyi* green fluorescent protein of the present invention were changed based on human codon bias.

The encoded protein of the present invention, when expressed in mammalian cell lines gives strong green fluorescence.

The synthetic DNA of the present invention can be used as a fluorescent tag for monitoring the activities of fusion partners using known image based techniques.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides a synthetic cDNA, based on the orange seapen *Ptilosarcus gurneyi* green fluorescent protein sequence, to encode a green fluorescent protein in which the majority of the codons are the favored or the most favored codons for mammalian expression systems. This process of codon preference modification when going from a native species to a host species, especially of going into a human cell-line host, is called "humanization".

A humanized gene is one that has been adapted for expression in human cells by replacing at least one, and most preferably, a significant number of the codons in the native gene codons with codons that are most frequently used in human gene expression. Thus, the native codon usage is replaced with a codon that is more favorable for

translation in a human or mammalian cell line. One reason for low expression of foreign genes in mammalian expression systems is the poor translation efficiency of the mRNA in the mammalian, and especially human cell environment. The reason for this is the difference in abundance of particular isoacceptor tRNA's that are different in human cells than those found in other organisms. In this instance the isoacceptor tRNA's are different in the *Ptilosarcus gurneyi* orange seapen than in human cells. Making the codon usage in the foreign gene match the prevalent isoacceptor tRNA subpopulation leads to improved translation efficiency, thus improved expression of the foreign gene in human cells.

The use of codon preference modification at the cDNA level results in higher levels of expression of the modified DNA molecule. Higher levels of expression leads to higher protein yield thus higher fluorescent signal in mammalian cells expressing the modified cDNA. The encoded protein of the present invention has 239 amino acid residues. In comparison to the wild type *Ptilosarcus gurneyi* green fluorescent protein (238 amino acid residues), codons for 145 amino acids were changed based on human codon preferences. One amino acid (valine) was added at the amino terminus to be the second amino acid residue in this protein. The encoded protein, when expressed in mammalian cell lines, gives strong green fluorescence. Generally, the green fluorescence to be achieved by the present invention is the production of light visible to the naked eye for qualitative purposes. Thus, the amount of the component of the bioluminescence reaction need not be stringently determined or met. It must be sufficient to produce light. The synthetic DNA of the present invention can be used as a fluorescent tag for monitoring the activities of its fusion partner, as described herein, using known imaging based approaches.

As will be appreciated by those skilled in the art, gene synthesis is performed by piecing together small pieces of double-stranded synthetic DNA. Each small double-stranded synthetic DNA is pieced together with even smaller single-stranded oligonucleotides. To place together one piece of the double-stranded DNA, 4 oligonucleotides are required. Figure 18 shows oligo 1 and oligo 2 are complimentary to each other in part of them. Thus, they can anneal to each other and can be extended by a DNA polymerase, such as for example Taq polymerase. The extended template then works as the template for an amplification reaction (using, for example. Taq polymerase)

employing oligo 3 and oligo 4 as the sense and anti-sense primers, respectively. The extension of the template and the amplification reaction are actually performed at the same time in the same test tube, and thus no separate step is shown in Figure 18. Two pieces of double-stranded synthetic DNA, from a synthesis scheme as for example the synthesis scheme set forth in Figure 18, may serve as the template for a new round of synthesis as long as they contain an overlapping region that can anneal to each other. This process may be repeated several times in order to create a long synthetic gene that can not be synthesized in one step. See J Hass et al., Codon usage limitation in the expression of HIV-1 envelope glycoprotein, Curr. Biol., Vol 6 (3), pages 315-324 (March 1996).

The fluorescent protein encoded by the modified humanized cDNA of the present invention is substantially identical to the wild type *Ptilosarcus gurneyi* fluorescent protein at the amino acid level, with the exception that the present invention provides for the addition of a single valine residue at position number 2 from the amino terminus. The absorption and emission spectra of the hPtFP in COS-1 cells was unchanged as compared to the wildtype *Ptilosarcus gurneyi.*

Figure 10 shows the double stranded nucleotide sequence (bottom two rows) of the entire coding region and the deduced amino acid sequence (top row) of the humanized *Ptilosarcus gurneyi* fluorescent protein (hereinafter "hPtFP") of the present invention, as well as the start and stop codon sequence, untranslated regions and restriction sites. Figure 11 shows the full length coding sequence of hPtFP of the present invention including regions upstream and downstream to the coding region.

Figure 12 shows the full length coding sequence of the hPtFP of the present invention from start codon to stop codon. Thus the total length of the humanized *Ptilosarcus gurneyi* fluorescent protein (hPtFP) of the present invention is 239 amino acid residues versus 238 for the wild type *Ptilosarcus gurneyi* fluorescent protein (PtFP). It will be appreciated that at the nucleotide level, approximately 61% of the wild type codons have been changed based on human codon bias. Figure 15 shows the codon usage table for human system, compiled from 22747 CDS's (10965560 codons) based on GenBank Release 118.0 (June 15, 2000), obtained from Kazusa DNA Research Institute

(Japan). Figure 15 shows the following fields: [triplet] [amino acid] [fraction, % of gene using the particular codon] [number of codons examined].

Figure 9 shows the wild type *Ptilosarcus gurneyi* (PtFP) nucleotide sequence (top row) compared to the nucleotide sequence (bottom row) encoding the hPtFP of the present invention. The difference at the nucleotide level (versus codon level) is that the hPtFP open reading frame (including the stop codon) contains 720 nucleotides, whereas the PtFP open reading frame (including the stop codon) contains 717 nucleotides. It will be appreciated by those skilled in the art, that without counting the extra valine introduced into hPtFP of the present invention, there are 166 nucleotide differences between the 717 nucleotides compared, or 23.15 percent difference (or 76.85% identity). If the stop codon is excluded in this comparison, there are 165 nucleotide differences in the 714 nucleotides compared, or 23.10 percent difference (or 76.9% identity). The synthetic DNA of the present invention was subcloned into an expression vector M2 (Cellomics, Inc., Pittsburgh, PA, USA) after restriction digestion of both DNAs with HindIII (New England Biolabs, Inc., Beverly, MA, USA) and NotI (New England Biolabs, Inc., Beverly, MA, USA) restriction endonucleases. The resulting expression vector was then used to transfect COS1 cells (CRL-1650, American Type Culture Collection [ATCC], Manassas, Virginia, USA) using FUGENE 6 transfection reagent (Roche Molecular Biochemicals, Indianapolis, IN, USA) following the protocol supplied by the manufacturer. Figure 16 shows an restriction endonuclease cleavage map of M2. Figure 17 shows the coding sequence of M2. M2 is a derivative of pCI-neo (Promega, Madison, WI, USA). Forty-eight (48) hours post transfection, the fluorescence was observed with an inverted epi-fluorescent microscope using a filter set for observing fluorescence as set forth in Figure 1. Figure 1 shows the expression of hPtFP in COS1 cells transiently transfected with the synthetic DNA of the present invention (right panel). The cells were counter stained with Hoechst 33342 (Molecular Probes, Eugene, OR, USA), a nuclear stain, Figure 1 (left panel).

Forty-eight hours after initial transfection with the synthetic green fluorescent protein DNA of the present invention, COS1 cells were trypsinized and were kept in suspension. The absorption and emission spectra of the live cells expressing the hPtFP of the present invention were then measured, as shown in Figure 2. Figure 2 shows the in

situ fluorescence of the humanized *Ptilosarcus gurneyi* fluorescent protein (hPtFP) in COS1 cells.

To compare the expression of wild type *Ptilosarcus gurneyi* fluorescent protein DNA with humanized *Ptilosarcus gurneyi* green fluorescent protein synthetic DNA as described above, an expression vector for the wild type *Ptilosarcus gurneyi* DNA was constructed by cloning the wild type PtFP into the expression vector M2, and thus, these two expression vectors under comparison differed only in their coding regions. Both DNA constructs were purified using QIAGEN plasmid kit (QIAGEN Inc., Valencia, CA, USA) folowing the instructions supplied by the manufacturer. The purified DNA preps were quantitated by reading the optic absorption at 260 nm (nanometers) with a HP8453 UV- visible spectrophotometer (Agilent Technologies, Palo Alto, CA, USA) and calculated based on 1 O.D. = 50 ng (nanograms) DNA. An identical amount of wild type *Ptilosarcus gurneyi* fluorescent protein DNA and the hPtFP was used to transfect COS1 cells, respectively, under identical conditions using FUGENE6 reagent, as described above. Forty hours post-transfection cells were fixed with 3.7 percent formaldehyde in the presence of 10 micrograms/milliliter of Hoechst 33342 (Molecular Probes, Eugene, OR, USA). The fluorescent images of the cells were then acquired using ARRAYSCAN II instrument (Cellomics, Inc., Pittsburgh, PA, USA) with 10X objective and filter setting at "FITC broad" (excitation = 365+/- 25nm, emission = 450 +/- 30 nm for Hoechst 33342- for fluorescent stain of nuclei, and excitation = 475 +/- 20nm, emission = 535 +/- 22.5 nm for hPtFP). U. S. Patent No. 5,989,835 describes the ARRAYSCAN II optical system and is incorporated by reference herein. The acquired images were then analyzed using a desktop client of ARRAYSCAN II instrument by identifying the nuclear region, and the intensity of the hPtFP was measured in the identified nuclear area. Figure 3 shows the comparative fluorescence measurements of COS1 cells transfected with the humanized cDNA of the present invention and the wild type cDNA. Figure 3 shows that the synthetic *Ptilosarcus gurneyi* green fluorescent protein DNA of the present invention emits stronger fluorescent signals than the wild type *Ptilosarcus gurneyi* green fluorescent protein. The results shown in Figure 3 confirm the Applicant's visual observation (qualitative) that the hPtFP DNA of the present invention produces more

hPtFP expressing cells and brighter cells than the wild type "PtFP" DNA in transient transfection.

The humanized synthetic DNA does not have toxic effects on the host cells, which aids in its increased stable expression. Stable expression of the hPtFP of the present invention was achieved in HEK293 (CRL-1573 ATCC, Manassas, VA, USA) and A549 (CCL-185, ATCC, Manassas, VA, USA) cell lines. The cells were co-transfected with the hPtFP construct and pSV2-neo (Cat # 37149, ATCC, Manassas, VA, USA) with FUGENE 6 transfection reagent (Roche Molecular Biochemicals, Indianapolis, IN, USA) following the manufacturer's instructions. Two days after transfection, cells were treated with 0.4 mg/ml (milligram/milliliter) G418 in normal growth medium. After treating with G418 for two weeks, drug resistant cells were isolated or pooled. A mixture of stably transfected HEK293 mixed population (in which about 30% of the cells expressing hPtFP of this invention) were plated out in 96 cell micro plates. After 2, 4, or 6 days incubation, the cells were fixed with 3.7% formaldehyde for 20 minutes at room temperature (25° Centigrade). The percentage of positive cells was quantitated with ARRAYSCAN II instrument, as described herein.

Figure 4, left panel, shows stably transfected HEK293 cells expressing the hPtFP of the present invention, and the right panel shows stably transfected A-549 cells. To compare the growth rates of cells expressing or not expressing hPtFP, a mixture of stably transfected HEK293 mixed population (in which about 30% of the cells expressing hPtFP of this invention) were plated out in 96-well micro plates. After 2, 4, or 6 days incubation, the cells were fixed with 3.7% (percent) formaldehyde for twenty minutes at room temperature (25 degrees Centigrade). The percentage of positive cells was quantitated with ARRAYSCAN II instrument, as described herein. Figure 5 shows the quantitation of positive cells from a mixture of stably transfected HEK293 cell population wherein approximately thirty percent (30%) of the cells expressing hPtFP were plated and incubated as described above. Figure 5 shows that the percentage of positive cells did not change during cell passage under no selection, indicating that the expression of the humanized *Ptilosarcus gurneyi* fluorescent protein of this invention is not toxic to the cells.

The hPtFP of this invention is useful as a fusion partner for tagging purposes. The hPtFP of this invention was fused to a model type-I single span transmembrane protein, human CD7 linked to the reactant target domain of the C-terminus of CD7. Human CD7 is a type-I single span transmembrane protein. CD7 is a member of the immunoglobulin gene superfamily well known by those skilled in the art and is a reliable clinical marker of T-cell acute lymphocytic leukemia. The fusion protein was expressed in COS1 cells by transient transfection using FUGENE 6 reagent, described hereinbefore, following the protocol supplied by the manufacturer. The distribution of the chimeric protein is similar to CD7 (no fusion partner) when transiently expressed by COS1 cells.

Figure 6, Panel A, shows C0S1 cells transiently transfected with human CD7 and stained with a monoclonal antibody against human CD7 (CD7 Ab-2, clone 124-1D1, Labvision Corp., Freemont, CA, USA). Figure 6, Panels B and C show COS1 cells transiently transfected with CD7-hPtFP of this invention. Panel B shows staining using a monoclonal antibody against human CD7 and Panel C shows a direct observation of the CD7-hPtFP fusion protein. The CD7-hPtFP chimera exhibits comparable localization as the untagged CD7.

The hPtFP is also useful in constructing biosensor systems. For example, the hPtFP may be used to construct protease biosensors for which the basic principle of the protease biosensors is to spatially separate the reactants from the products generated during a proteolytic reaction. The separation of products from reactants occurs upon proteolytic cleavage of the protease recognition site within the biosensor, allowing the products to bind to, diffuse into, or be imported into compartments of the cell different from those of the reactant. This spatial separation provides a means of quantitating a proteolytic process directly in living or fixed cells. A design of the biosensor provides a means of restricting the reactant (uncleaved biosensor) to a particular compartment by a protein sequence ("reactant target sequence") that binds to or imports the biosensor into a compartment of the cell. These compartments include, but are not limited to any cellular substructure, macromolecular cellular component, membrane-limited organelles, or the extra-cellular space. Given that the characteristics of the proteolytic reaction are related to product concentration divided by the reactant concentration, the spatial separation of

products and reactants provides a means of uniquely quantitating products and reactants in single cells, allowing a more direct measure of proteolytic activity.

The molecular based biosensors may be introduced into cells via transfection and the expressed chimeric proteins analyzed in transiently transfected cell populations or stable cell lines. They may also be pre-formed, for example by production in a prokaryotic or eukaryotic expression system, and the purified protein introduced into the cell via a number of physical mechanisms including, such as for example, but not limited to, micro-injection, scrape loading, electroporation, and signal-sequence mediated loading, etc.

Measurement modes may include, such as for example, but are not limited to, the ratio or difference in fluorescence, luminescence, or phosphorescence: (a) intensity; (b) polarization; or (c) lifetime, between reactant and product. These latter modes require appropriate spectroscopic differences between products and reactants. For example, cleaving a reactant containing a limited-mobile signal into a very small translocating component and a relatively large non-translocating component may be detected by polarization. Alternatively, significantly different emission lifetimes between reactants and products allow detection in imaging and non-imaging modes.

One example of a family of enzymes for which this biosensor can be constructed to report activity is the caspase family. Caspases are a class of proteins that catalyze proteolytic cleavage of a wide variety of targets during apoptosis. Following initiation of apoptosis, the Class II "downstream" caspases are activated and are the point of no return in the pathway leading to cell death, resulting in cleavage of downstream target proteins. Specifically, the biosensors described herein are engineered to use nuclear translocation of cleaved hPtFP as a measurable indicator of caspase activation. Additionally, the use of specific recognition sequences that incorporate surrounding amino acids involved in secondary structure formation in naturally occurring proteins may increase the specificity and sensitivity of this class of biosensor.

The protein biosensors herein disclosed can be adapted to report the activity of any member of the caspase family of proteases, as well as any other protease, by a substitution of the appropriate protease recognition site in any of the constructs. These biosensors can be used to detect in vivo activation of enzymatic activity and to identify

specific activity based on cleavage of a known recognition motif. This screen can be used for both live cell and fixed end-point assays, and can be combined with additional measurements to provide a multi-parameter assay, as is well known in the art.

Thus, another aspect of the present invention provides recombinant nucleic acids encoding a protease biosensor, comprising: (a) a first nucleic acid sequence encoding a *Ptilosarcus gurneyi* green fluorescent protein having its codon usage optimized for expression in human cells that encodes at least one detectable polypeptide signal; (b) a second nucleic acid sequence that encodes at least one protease recognition site, wherein the second nucleic acid sequence is operatively linked to the first nucleic acid sequence that encodes at least one detectable polypeptide signal; and (c) a third nucleic acid sequence that encodes at least one reactant target sequence, wherein the third nucleic acid sequence is operatively linked to the second nucleic acid sequence that encodes at least one protease recognition site.

Generally, a protease biosensor is composed of multiple domains, including at least a first detectable polypeptide signal domain, at least one reactant target domain, and at least one protease recognition domain, wherein the detectable signal domain and the reactant target domain are separated by the protease recognition domain. Thus, the exact order is not generally critical as long as the protease recognition domain separates the reactant target and first detectable signal domain. For each domain, one or more of the specified sequences is present.

The organizations of the biosensors are shown in Figure 7 (Caspase 3) and Figure 8 (Caspase 8). Those persons skilled in the art will recognize that any one of a wide variety of protease recognition sites, reactant target sequences, polypeptide signals, and/or product target sequences can be used in various combinations in the protein biosensor of the present invention, by substituting the appropriate coding sequences into the multi-domain construct. Non-limiting examples of such alternative sequences are shown in Figures 7 and 8. Similarly, those skilled in the art will recognize that modifications, substitutions, and deletions can be made to the coding sequences and the amino acid sequences of each individual domain within the biosensor, while retaining the function of the domain. Such various combinations of domains and modifications,

substitutions and deletions to individual domains are within the scope of the instant invention.

As used herein, the term "coding sequence" or a sequence which "encodes" a particular polypeptide sequence, refers to a nucleic acid sequence which is transcribed (in the case of DNA) and translated (in the case of mRNA) into a polypetide in vitro or in vivo when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. A coding sequence can include, such as for example, but is not limited to, cDNA from prokaryotic or eukaryotic mRNA, genomic DNA sequences from prokaryotic or eukaryotic DNA, and synthetic DNA sequences. A transcription termination sequence will usually be located 3' to the coding sequence.

As used herein, the term DNA " control sequences" refers collectively to promoter sequences, ribosome binding sites, polyadenylation signals, transcription termination sequences, upstream regulatory domains, enhancers, and the like, which collectively provide for the transcription and translation of a coding sequence in a host cell. Not all of these control sequences need always be present in a recombinant vector so long as the DNA sequence of interest is capable of being transcribed and translated appropriately.

As used herein, the term "operatively linked" refers to an arrangement of elements wherein the components so described are configured so as to perform their usual function. Thus, control sequences operatively linked to a coding sequence are capable of effecting the expression of the coding sequence. The control sequences need not be contiguous with the coding sequence, so long as they function to direct expression thereof. Thus, for example, intervening untranslated yet transcribed sequences can be present between a promoter sequence and the coding sequence and the promoter sequence can still be considered "operatively linked" to the coding sequence.

Furthermore, a nucleic acid coding sequence is operatively linked to another nucleic acid coding sequence when the coding region for both nucleic acid molecules are capable of expression in the same reading frame. The nucleic acid sequences need not be contiguous, so long as they are capable of expression in the same reading frame. Thus,

for example, intervening coding sequences, and the specified nucleic acid coding regions can still be considered "operatively linked".

The intervening coding sequences between the various domains of the biosensors can be of any length so long as the function of each domain is retained. Generally, this requires that the two dimensional and three-dimensional structure of the intervening protein sequence does not preclude the binding or interaction requirements of the domains of the biosensor, such as product or reactant targeting, binding of the protease of interest to the biosensor, fluorescence or luminescence of the detectable polypeptide signal, or binding of fluorescently labeled epitope-specific antibodies.

Within this application, unless otherwise noted, the techniques utilized may be found in any of several well-known references such as Molecular Cloning: A Laboratory Manual (Sambrook, et al. 1989, Cold Spring Harbor Laboratory Press), Gene Expression Technology (Methods in Enzymology, Vol. 185 edited by D. Goeddel , 1991, Academic Press, San Diego, CA), "Guide to Protein Purification" in Methods in Enzymology (M.P. Deutscher, ed., (1990) Academic Press, Inc.); PCR Protocols: A Guide to Methods and Applications (Innis, et al. 1990. Academic Press, San Diego, CA), Culture of animal Cells: A Manual of Basic Technique, 2nd Ed. (R.I. Freshney, 1987. Liss, Inc. New York, NY) Gene Transfer and Expression Protocols, pp. 109-128, ed. E.J. Murray, The Human Press Inc. Clifton, NJ), and the Ambion Catalog (Ambion, Austin, TX).

The biosensors of the present invention are constructed and used to transfect host cells using standard techniques in the molecular biological arts. Any number of such techniques, all of which are within the scope of this invention, can be used to generate protease biosensor-encoding DNA constructs and genetically transfected host cells expressing the biosensors. The biosensors disclosed in pending published patent application WO 0026408, entitled "A System For Cell Based Screening" provide examples of such biosensors; PCT/US99/25431 is made of record and incorporated by reference into this patent application. The non-limiting examples that follow demonstrate one such technique for constructing the biosensors of the invention. For example, by changing the protease recognition sequence of the sensors shown herein into the recognition sequence for other caspases or other intracellular proteases, such as for

example calpain and cathepsins, new specific protease sensors can easily be generated. Other examples of green fluorescent protein-based biosensors include, but are not limited to, fluorescent resonance energy transfer (FRET) based, green fluorescent protein-based caspase sensor disclosed by J. Jones et al., J. Biomol. Screen., Vol. 5 (5), pages 307-318 (Oct. 2000), A. Miyawaki et al., Nature, Vol. 388 (6645), pages 882-887 (Aug. 1997), and J.P. Waud et al., J. Biochem., Vol. 357 (Pt. 3), pages 687-697 (Aug. 2001).

In addition to the full length coding sequence hPtFP of the present invention as shown in Seq. ID No. 1 several truncation mutants are disclosed ranging from truncations at the 5' (amino) terminus to truncations at the 3' (carboxy) terminus. SEQ ID No. 2 shows the amino acid sequence of the full length hPtFP of the present invention. SEQ ID No. 3 shows a truncation mutant of the present invention wherein the truncation occurs at the 3' (carboxy) terminus, specifically, including amino acid sequence 1-224. SEQ ID No. 4 shows a truncation mutant of the present invention wherein the truncation occurs at the 5' (amino) terminus, specifically including amino acid sequence 10-229. Figure 13 shows deletion mutants of the hPtFP of the present invention that were constructed. The fluorescent intensity upon visual inspection of each construct is shown in Fig. 13. The deletion mutants of the hPtFP of the present invention were constructed and transiently transfected into HeLa cells. The deletion mutants were created by employing PCR (polymerase chain reaction) technology, as known by those skilled in the art, and were sub-cloned into the expression vector M2 in which the expression in the mammalian systems is driven by the CMV (cytomegalovirus) promoter. All mutants and the full length hPtFP of this invention, expression constructs were designed to be identical in the non-coding region. All coding regions constructed for this comparison as shown in Fig. 13 retain M (methionine) and V (valine) as the first and second amino acids, respectively. The plasmids were then transfected into Hela cells and observed for fluorescence 24 hours after transfection as shown in Fig. 13. It will be appreciated by those skilled in the art that the truncation mutants of the present invention may be employed as fluorescent tags for monitoring the activities of its fusion partners using an image based approach as a biosensor.

Figure 14 shows HeLa cells (CCL-2, ATCC, 10801, Manassas, VA, USA) transfected with with the hPtFP-Caspase-8 biosensor with FUGENE 6 reagent (Roche

Molecular Biochemicals, Indianapolis, IN, USA). Twenty four hours after transfection, the HeLa cells were treated with staurosporine (Sigma-Aldrich, St. Louis, MO, USA), at 1nM (nano molar) or 10nM. Fluorescent signals from the cells were observed at the 6 hours and 24 hours, respectively, after addition of the staurosporine to the medium.

Whereas particular embodiments of this invention have been described herein for purposes of illustration, it will be evident to those persons skilled in the art that numerous variations of the details of the present invention may be made without departing from the invention as defined in the appended claims that follow the SEQUENCE LISTING.